Reflection learning with Neural Network Ensemble

Mars Gao
gao463@purdue.edu

Reflection learning with Collaborative Neural Network Groups

Liyao Gao
Department of Computer Science
Purdue University
gao463@purdue.edu

Abstract

For the current architecture of neural networks, it usually requires a high training cost in time and computation. From our perspective, the current methods in deep learning might not be optimal in architecture and it fails to have an efficient learning strategy. To solve these problems, in this paper, we would like to introduce the Collaborative Neural Network Group (CNNG). CNNG is a series of neural networks that work collaboratively to handle different tasks separately in the same learning system. It is evolved from a single neural network by our designed algorithm — Reflection. In this way, based on different situations extracted by the algorithm, the CNNG is able to perform different strategies when predicting the input data. In our implementation, the CNNG is combined by several relatively small neural networks. We provide a series of experiments to evaluate the performance of CNNG compared to other learning methods on three public datasets. The CNNG is able to get a higher accuracy with a much lower training cost. With CNNG by reflection, we can reduce the error rate 74.6% by average and reach a high accuracy for many tasks, which is superior to VGG and ResNet on the tested datasets. For Fashion-MNIST and EMNIST, it can reach 98.81% and 90.88% which is the best performance currently. Moreover, the required training time is usually less than 40 minutes in our experiments. Details can be found in the experiment part.

Introduction

Recently, a series of recent designs in the deep neural network has been reported to reach a satisfying performance in different tasks, including image recognition, speech recognition, and machine translation (LeCun, Bengio, and Hinton 2015). However, still, there are lots of problems in deep learning behind the high accuracy. The main problem is the high training cost. As far as we are concerned, the current deep learning often requires a huge amount of resource in time and computation when performing a learning task. The architecture requires to have a large number of layers to guarantee a good performance. There exists a belief that a very deep neural network is able to solve any learning task. It results in the current trend that it needs a much higher expansion in model size for a large-scale learning task, such as the use of VGG. At the same time, from the aspect of computation and optimization, the training will become much more difficult with an oversized network. Since the model is becoming larger, the computation has been more complex. Also, as the number of parameters goes up, it is much harder for an optimization method to find the optimal solution.

More specifically, there can be two concerns in this paper. Firstly, the current architecture of deep learning is not entirely biologically reasonable. The neural networks are believed to be able to solve cognitive tasks like face recognition and machine translation because it is believed to be a biological simulation. However, it is hard to claim that the current architecture of the neural network is a satisfying simulation. A point that we observe here is: in the human brain, different tasks are handled by different neural networks. For example, the audio area will be active when listening, the visual area will be active when performing visual recognition and Wernicke's area will be active when performing language comprehension. It is nonsense to use the Wernicke's area to perform visual understanding. A proof is that if a certain area in the brain was wounded, the related functions will be affected. Secondly, the strategy of human learning is missing in the current training process. When a human is performing a learning task, the learning strategies play an important role. The learning strategies can largely influence the performance of the study, for both accuracy and efficiency (Riding and Sadler-Smith 1997). From the perspective of learning science, the use of learning strategies can help to utilize the benefit of the learning architecture in the brain. For the same idea in artificial neural networks, the missing of appropriate learning strategies might hinder the opportunity to utilize the benefit of the architecture of the learning system. The current learning process might also affect the performance.

In this paper, we would like to introduce the Collaborative Neural Network Group (CNNG) by reflection. Collaborative Neural Network Group is an architecture that combined with a series of neural networks. Reflection is the learning algorithm that is able to generate the CNNG from a single neural network. It is originated from the learning strategy of humans (Boud, Keogh, and Walker 2013). In a learning task, human will perform a reflection that reconsiders the problem and analyze their mistakes. As a similar approach in a neural network way, a general neural network will be

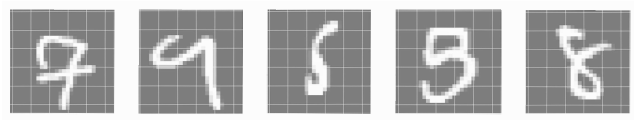

Figure 1: Typical error cases for neural network in MNIST

initially trained for a learning task. Then, the error cases of the general network will be collected. The algorithm will classify the error cases into different clusters and initialize a corresponding number of neural networks to be trained by the error clusters. The networks that focus on the different error cases will become the specialist neural network. For the last step, a task classifier will be trained based on the error clusters to determine which network to use for an incoming data. This is the way how a single neural network will be evolved into the CNNG by reflection. The networks in CNNG are viewed to be different strategies to use when processing the tasks.

From our perspective, the CNNG by reflection can be a better method to use for a large-scale learning task. Firstly, it might lower the difficulty of training. In the CNNG, the ambiguous cases are separated into different clusters and will be processed by different networks. Therefore, it could help to decrease the difficulty of training. Also, after reflection, the specialist networks will be used on their specified tasks. It is able to have a better performance compared to the general network on the assigned tasks. In this way, for every time that the task classifier is making a correct assignment, the CNNG will have a higher probability to predict the output correctly. The using of the specialist network will not only help the CNNG to use a network with higher accuracy on the task, but also remove the cases that originally is ambiguous for the general network. Therefore, the accuracy can be greatly improved. From this point, we believe that CNNG by reflection can be a better method to solve the problem of large-scale learning tasks. We view this as an important method to solve large-scale learning tasks with a higher efficiency and accuracy.

A series of experiments has been provided in this paper. Based on the experiments, we observe the superiority of CNNG in both training time and accuracy compared to the current methods. The training time is less than an hour for all the experiments without GPU acceleration. For the accuracy, we compared the performance of CNNG by reflection with other typical learning methods. We reached the highest reported accuracy in both Fashion-MNIST and EMNIST datasets. We tested the result thoroughly on two image datasets: MNIST, Fashion-MNIST and EMNIST. Our experiment reached a satisfying result and our model can largely lower the loss. Specifically for MNIST, we are able to reach the accuracy of 99.65% with three simple feedforward networks (3 layers) with one training epoch; it lowers the error rate by 71.9%. For Fashion-MNIST, we are able to lower the error rate by 66% and reach the accuracy of 98.81%. The highest reported accuracy for ResNet is 92% and the accuracy for VGG is 93.5%. For EMNIST, the CNNG after reflection is able to reduce the error rate by 60%, and we reached the accuracy of 90.88% with three simple feedforward networks (4 layers) with one training epoch. The highest reported accuracy of ResNet is 88.5%.

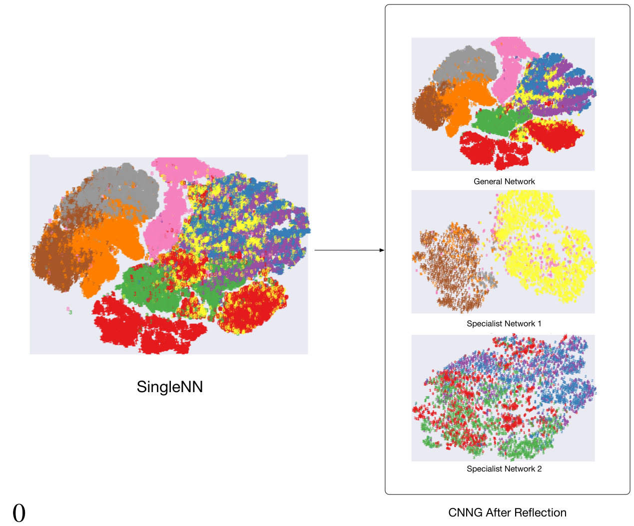

Figure 2: tSNE visualization of the data processed after Reflection with CNNG

Beside of the performance, we also provided a series of observations about this architecture to explain its mechanism. The performance improvement since the training difficulty has been reduced and the accuracy comes from the Specialist Networks. Firstly, the ability of the reflection that can lower the training difficulty has been proved by detailed evaluation of different networks. Figure 2 is the tSNE visualization of data in Fashion-MNIST. It shows the difference between using a single neural network to handle the data and using CNNG by reflection. Clearly, the data on the right is intuitively easier to be trained by a learning architecture since the cases are less ambiguous. This point can be further explained by the detailed evaluation on accuracy. The original corner cases that are difficult has been eliminated from the general situation; therefore, the accuracy of General Network on general situation can reach 99.41%, improved from 85% training with single neural network. The specialist networks handle the cases that originally will be falsely predicted by the General Network. As we can see in Figure 2, they are having less difficult learning tasks. According to the experiments, in the case of Fashion-MNIST, the accuracy of Specialist Network can reach the accuracy of 90.5% and 88.03%. Other experiments provide clues of choosing the correct number of specialist network and task classifier. Detailed data and experiment setup can be found in the evaluation part.

In summary, our contribution in this paper can be summarized as follows:

1. We design Collaborative Neural Network Group by Reflection as a new deep neural network architecture to solve the learning tasks, which inspired by the architecture of the brain and human learning strategy.

2. We implement the CNNG architecture with task clas-

sifier and neural network groups, and use the decision tree algorithm combined with K-Means to perform Reflection as a simulation to human learning strategy.

3. We conduct experiments, on the tested datasets, to show its superiority to lots of existing architecture, such as VGG and ResNet, in accuracy and training cost. Also, experiments have been performed to show its mechanism to lower the training difficulty. It explains the reason of the increased accuracy.

Related works

Cortex and brain neural network

In the development of recent neuroscience and cognitive science, the architecture of cortex and association areas are mentioned a lot by the scientists (Guerguiev, Lillicrap, and Richards 2017) (Jane X Wang 2018) (Roelfsema and Holtmaat 2018) (Guangyu Robert Yang 2017). In these works, researchers try to provide better mathematical model to simulate a more complex architecture and mechanism in human brain to implement a better learning system. People view this as the next step of the cognitive system. When our brain is performing a cognitive task, the cortex will assign the input and activate different association area with the data. Different association area including visual, language, logical thinking and critical thinking are neural networks that specified on a different specific field of tasks (Kandel et al. 2000). Compared to a very deep neural network, our work can be more biological plausible. The reflection will create a set of neural networks that will be used in different situations. The task classifier can be viewed as cortex in the brain and the specialized neural networks can be viewed as the association areas.

Neural network ensembles

Neural network ensembles (Hansen and Salamon 1990) (Krogh and Vedelsby 1995) (Huang et al. 2017) are aimed to solve a large task with not only one neural network. The neural network ensembles is consist of a set of simple neural networks. All of the networks are participated for prediction every time. The final output is combined by the output of the networks by weights. This can be viewed as a similar method to Adaboost, a combining pattern classifier. An interesting research in this area is the Cooperative Neural Network Ensembles (Islam, Yao, and Murase 2003). The Cooperative Neural Network Ensembles is different from the neural network ensembles. Another similar approach is Cooperative Ensembles Learning System (CELS) (Liu and Yao 1999). For the CELS method, it is trying to divide the datasets into several parts. Then, using the separated data to train with different neural networks. We believe that this method might increase the difficulty of training. Every time after the split of data, the individual networks are having less amount of training data. For every time of division, it should try to guarantee that the subtasks are becoming less complex. However, this could be an egg-chicken problem which is nearly impossible, and it was not thoroughly discussed in the CELS. The CNNG with reflection might provide another approach to improve the performance when avoiding this problem.

Reflection

Design Motivation

In general, the motivation of this architecture came from the aspect of cognitive science, biology and learning science. The motivation of our design will be described in detail as follows:

Solve different problems with different networks.

As we stated before, one of the difference currently between artificial neural network and a real human brain is that human brain is highly differentiated. There exist different areas to handle different tasks. In this paper, we propose the design of Collaborative neural network groups as an approach to implement this idea. For a learning task, the CNNG will try to define an initial pattern of the data. Data with different initial pattern will be assigned into different 'specialized' networks; then the CNNG will use different networks to recognize the main pattern, predicting the label of the input data. To explain the initial pattern, take MNIST as an example, the cases that we captured in Figure 1 can be highly ambiguous. It is hard for a neural network to learn from a learning task with a large amount of cases that have similar input but completely different output. The idea of initial pattern may help to avoid this situation by splitting the ambiguous cases at the first step. It is automatically determined by a decision tree but it can relate to underlying information such as nationality, culture, and education in the case of MNIST. Each network in the neural network groups could be specialized as different task and should be used for different situation.

Use reflection as a learning strategy to train a learning system.

Reflection is an important strategy in human learning. It is originated from the learning strategy of humans (Boud, Keogh, and Walker 2013). In a learning task, human will perform a reflection that reconsiders the problem and analyze their mistakes. More specifically, human will normally classify their mistakes into different situations and design different strategies to solve them. Then, the strategies will be retrained by the mistakes. This could be summarized as the reflective learning process of humans. After the reflection process, for an incoming task, human will first try to quickly overview the problem and decide which strategy to use; then process the task with the corresponding solution. In this way, humans could improve their performance on specific tasks using a set of strategies. From the view of a learning system, it has no information about the data before training process, so it is hard to define what is the initial patterns. This is the reason why we apply reflection here: it can be viewed as a step to gradually analyze and understand the whole datasets. Some clear initial pattern can be extracted based on correct and wrong predictions.

CNNG Model description

CNNG is a group of neural networks that will be used to process the input data. It is evolved from a single neural network by the reflection algorithm. The benefit of CNNG is the collaboration of several small networks which are specified on different special tasks individually can help to largely improve the accuracy. It requires a much lower training cost at

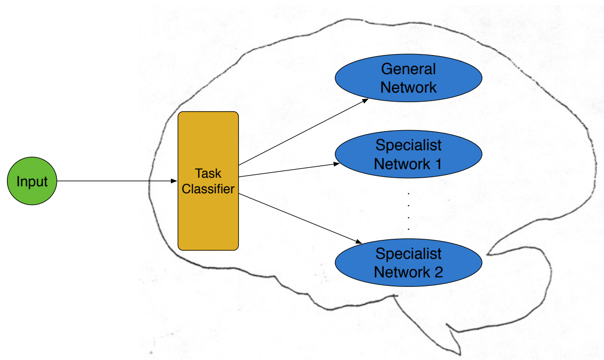

Figure 3: Architecture of CNNG

the same time. The architecture of the CNNG is combined with task classifier, general neural network and specialist neural networks.

Task classifier The task classifier is used for the CNNG to decide which is the best network to use when predicting the label of the input data. In our approach, the task classifier is a decision tree. The task classifier allows the neural networks to collaboratively work together as a group.

General neural network The general neural network is the network that is initially trained by all the training data. This will serve as a general situation when handling the input data. A reflection will be performed based on the general network. The error cases will be used to train the specialist neural networks.

Specialist neural network The specialist neural network is the neural network that will focus on different subtasks. It will be trained from the error that produced by the general network. It is viewed as the specialist in the system that will process the corner cases for general network.

For the prediction process of CNNG, the task classifier will first determine the best network to use for an input data. Then, the best network will take charge of the input data and predict the label.

Reflection Algorithm

The reflection algorithm here, in Algorithm 1, is motivated by learning strategy of humans. Humans frequently perform a reflection when they are approaching their extreme in performance in a learning task. Reflection is an important process in learning (Boud, Keogh, and Walker 2013). Normally, as we stated before, human will try to analyze the problem, divide them into different cases, and try to apply different methods to solve them separately. Based on this idea, we designed the concept of reflection on Collaborative Neural Network Group. Basically, the general neural network that will be trained initially to handle the general situation. This is served as the general training process. For the cases that still have a high error for the general network, it will be divided into different clusters. Then, a series of new neural networks will be initialized and trained with the input with a high error for the general neural network. This is served as the second training process. For the last step, the task classifier will be trained with inputs and labels of the network id. The task classifier is going to decide which network to use for input. We use the K-Means method here to decompose the error cases.

Number of specialist networks

Suppose k is the number of specialist network. The choice of k is important because it affects the hardness in training for the specialist network and task classifier here. The desired k here should try to *minimize the difficulty of the training of task classifier and specialist neural networks*. A kernel problem here is the estimation of the difficulty for a training set to be learned by a neural network. Specifically, on one hand, a very large k might not only involve in the problem of overfitting, but also may not be able to sufficiently train the specialist networks for limited amount of training data. The performance of task classifier might also be negatively affected for predicting more clusters. On the other hand, a relatively small choice of k might not simplify the problem. Therefore, the expected accuracy of the specialist network might fail to become better. The number k might need to be determined by the task in practice. In our implementation for MNIST and EMNIST, usually, we take $k = 2$. Experiments can be found in the evaluation part with different number of specialist networks.

Task Classifier Training

The task classifier plays an important role in reflection. It is served as a task assigner or a strategy decider in CNNG which will largely affect the performance. In the process of reflection, the input data are separated by features that determined by the error cases. The choice of task classifier should pair with the method we use in the reflection that is able to understand the patterns for the best. In our implementation, K-Means is used in splitting the error cases while reflecting. Because of the using of K-Means, the pattern of the label can be viewed as a group of constraints by numbers for x.

In this case, the decision tree method could possibly be the best classifier here. The decision tree uses a threshold to classify the data into different groups and this is the pattern that we would like the task classifier to learn. Based on our experiments, the decision tree method does have the best performance compared to other kinds of classical classifiers. Details can be found in the evaluation part.

Experiments and Evaluations

We applied the CNNG by reflection into different situations. We compared the general performance of CNNG with reflection compared to other methods. In this paper, we focused on the evaluation on the image tasks. We conducted detailed experiments on the performance of CNNG with reflection compared to SingleNN with different amount of training data, deeper neural networks. The evaluation of task classifier and the networks in CNNG individually can provide an insight to show where does the increase in performance come from.

The SingleNN that we use in the following experiments is a simple feedforward network. For the MNIST, the initial network contains three layers. For the Fashion-MNIST, the initial network contains four layers. For the EMNIST, the initial network contains four layers.

Algorithm 1 CNNG reflection algorithm

function REFLECT(X, Y, $ErrSet$)
    $SpecialistNetworkSet = []$
    $TaskClassifierDataDict = \{\}$
    $SpecialistsTrainingSet = Kmeans(ErrSet, SpecialistNum)$
    for all $trainingData \in SpecialistsTrainingSet$ **do**
        $network = TrainNetwork(trainingData)$
        $TaskClassifierDataDict.add(trainingData.getX())$
        $SpecialistNetworkSet.append(network)$
    end for
    $TaskClassifier = DecisionTree.train(TaskClassifierDataDict)$
    return $SpecialistNetworkSet, TaskClassifier$
end function

Task	CNNG Accuracy (*)	SingleNN Accuracy	CNN accuracy	Adaboost Accuracy	CNNE Accuracy
MNIST	**99.65%**	97.84%	96.35%	79%	98.6%
EMNIST	**90.88%**	82.32%	83.5%	65%	89.3%
Fashion-MNIST	**98.81%**	85.56%	91.9%	93.6%	92.22%

Table 1: Preformance of CNNG and SingleNN

Overall performance of CNNG to other methods

In this section, we provided an overall performance of CNNG and its comparison to other classic methods. The CNN here is combined with two convolutional layers and four linear layers. Adaboost is a statistical method of boosting neural networks. Four classifiers is used here. CNNE (Islam, Yao, and Murase 2003) is another incremental learning method which sequentially initiate a new network on the error. We evaluate the performance of CNNG by reflection with existing methods. For the CNNG here, it is combined with three simple feedforward networks. All of the following methods are trained with one epoch. We choose the type balanced in EMNIST, which contains 131000 images for 62 different outputs. The Adaboost method is using four classifiers.

Firstly, as we can see, the CNNG is having the best performance among other methods with low training effort. The test accuracy that it is able to reach is satisfying for 99.65% in MNIST, 98.81% in Fashion-MNIST, and 90.88% in EMNIST. It lowers the error rate by 48.4% in EMNIST, 91.8% in Fashion-MNIST and 82% for MNIST. Especially for the Fashion-MNIST and EMNIST, it has a higher accuracy compared to ResNet, which are 92% and 88.5%. It seems to be the best performance so far in these two datasets. Another point here is that the CNNG is training with a much smaller amount of time, which is less than 40 minutes for every tasks. Based on this experiment, we can see that CNNG by reflection is able to reach the highest performance than other methods with low training efforts.

Accuracy with different amount of training data

In this experiment, we aim to see a comparison between CNNG and singleNN when training with different amount of data. The following table shows its difference in performance with different amount of training data.

As we can observe in Figure 4, from the very beginning, the accuracy for CNNG by reflection is lower than SingleNN when only 5% data are used. The reason that the accuracy is low from the beginning is because of lack of training data. The specialist network only has several cases for training. But the performance of CNNG by reflection quickly goes up and reach a higher performance than SingleNN. We can see that the reflection can promise to improve the performance of the single neural network with a certain amount of training data. Based on this experiment, it is able to show that CNNG is able to reach a higher performance compared to the singleNN with a smaller amount of training data.

Comparison between Reflection and Deeper networks

In this experiment, we aim to test the performance between the performance of CNNG by reflection to simply making more layers of networks. Since compared to a SingleNN, the CNNG is having a larger size. Normally, a deeper neural network will have a better performance. A comparison between CNNG and a SingleNN with more layers can important for evaluating the CNNG. Our experiment aims to figure out the performance of CNNG by reflection and deeper networks. We did this experiment on MNIST, Fashion-MNIST and EMNIST.

Based on the result, as we can see in Figure 5, it is clear that the CNNG by reflection is superior to the networks with more layers. Firstly, when the neural network is adding layers, it is adding connections by 256 times. Even if the number of connections becomes larger, the performance is not improving as significant as CNNG, which only adding three times of connections in this implementation. Even if having

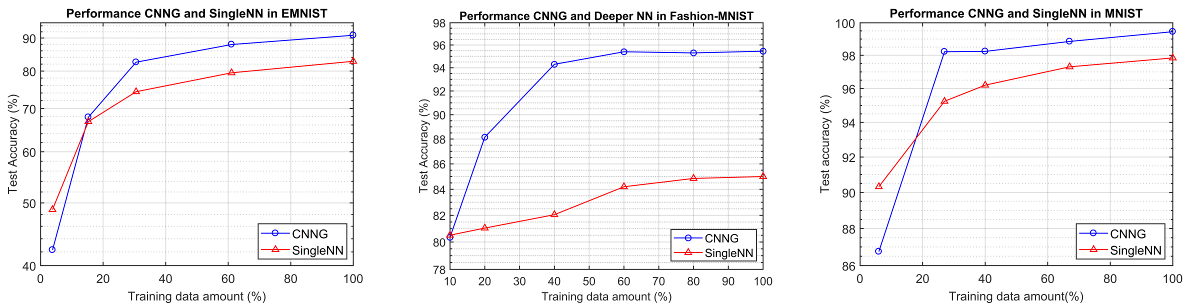

Figure 4: Different performing tasks by general network and specialist networks

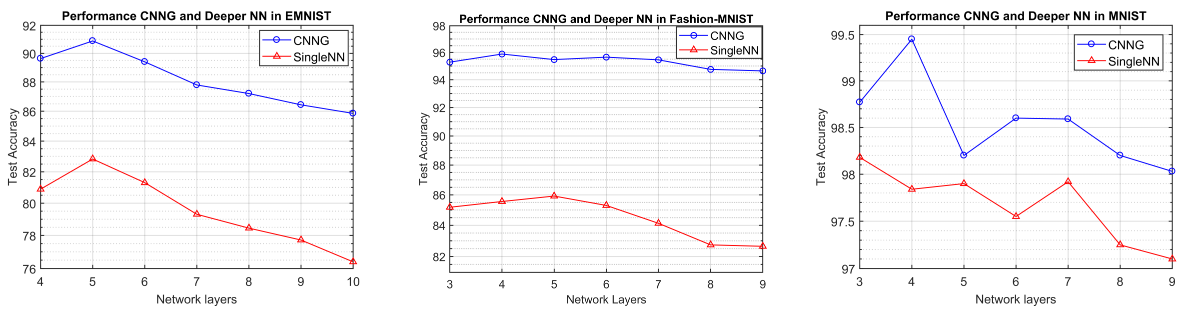

Figure 5: Comparison of CNNG to networks with more layers

nine or ten layers, the accuracy of CNNG reflecting on the single neural network is still higher than the best result. The CNNG after reflection is improving in a higher efficiency which has better performance with fewer connections. Secondly, making the network deeper might not guarantee the improvement in performance but the CNNG after reflection can make the accuracy better under the condition of limited amount of training data. As we can see, having four layers in MNIST, four layers in Fashion-MNIST and five layers in EMNIST, have the best performance among other models. The performance then goes down when adding more layers. A deeper network might require more data to train. But CNNG can still lead a promising improve accuracy. From this experiment, we believe that CNNG can more better in accuracy and efficiency compared to deeper networks.

Evaluation on the TaskClassifiers

The choice of TaskClassifier can largely affect the performance of the reflection. In this experiment, we provided a series of classical classification methods. We applied them in MNIST and EMNIST tasks as TaskClassifier. After the same training process, we evaluated the performance of different classical classification methods as the table following. In the following table, KNN stands for K-Nearest-Neighbour classifier and MLP stands for Multiple-Linear-Perceptron.

Based on the result, we can find out that both the decision tree method is the best. The performance of the general network here is 99.65% for MNIST, 98.81% for Fashion-MNIST and 90.88% for EMNIST. From the graph, we can see that the Naive Bayes is clearly not a probable choice here. The decision tree method is having the best performance among these classifiers. The KNN and MLP classifier is able to increase the performance in EMNIST task but for the MNIST task, the performance become even worse than the general network.

Different number of specialist networks evaluation

The selection of k is an important problem in this architecture. As we stated before, the k could affect the performance and there should exist a best k for a learning task. The selection of k should relate to the number of error cases, and the shape of the data. It should try to minimize the learning difficulty of training in task classifier and specialist networks. It is still impossible to find a method to determine the best k, so we provided a series of experiments here to show the accuracy with different k, in Figure 6.

In this figure, we can find out that when the number of specialist networks is increasing, the accuracy of CNNG goes up to a point and then goes down. There exist an optimal number of specialist network. In our opinion, as we stated before, the selection of k should help to reduce the training difficulty of task classifier, general network and specialist networks. When k is small, the error cases might still be highly ambiguous. So the accuracy goes up for the first. However, when k is large, since the number of error cases are limited, the specialist networks might fail to have enough training samples. It will result in a decrease for a large k.

Detailed evaluation of the neural networks individually in CNNG

This experiment aims to explain a detailed observation behind the improved performance in CNNG and provide evaluation

Task Datasets	Decision Tree (*)	KNN	MLP	Naive Bayes
MNIST	99.65%	96.32%	94.41%	36.78%
EMNIST	90.88%	83.80%	84.30%	20.70%
Fashion-MNIST	98.81%	96.12%	92.43%	32.47%

Table 2: Preformance of CNNG and SingleNN

Network	Percentage using by CNNG	Overall Accuracy	Accuracy on specified task
GeneralNet	72.5%	84.32%	99.43%
SpecNet1	12.4%	15.32%	96.96%
SpecNet2	14.9%	10.20%	93.33%

Table 3: Preformance of CNNG and SingleNN

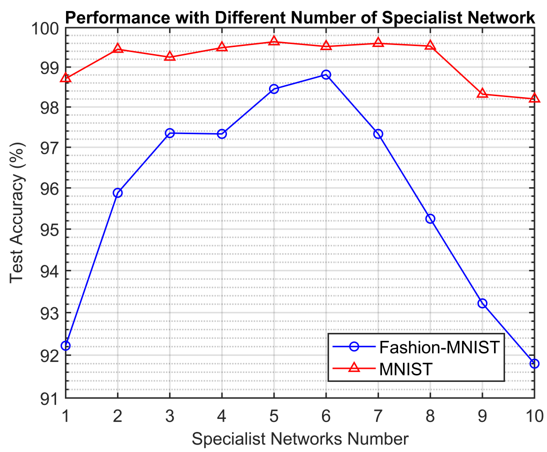

Figure 6: Performance of different number of specialist networks

of the reflection algorithm. The following shows the result of the networks in CNNG individually. It evaluates the accuracy on specified task is resulted by cross validation and the times that each network is used by CNNG in prediction. We also test the accuracy of all three networks of its overall performance on the whole task. We conducted this experiment using Fashion-MNIST.

According to this result, there are two points that we can summarize. Firstly, the reflection is successfully decomposing a large learning task into easier subtasks. More specifically, the specialist network is able to have a better performance in the specified tasks and the specified tasks have a lower difficulty in training. As we can see in the table, for the two specialist networks, they have a very low overall accuracy in overall, which are both around 10% to 20%. However, in their specified task, they have a much higher accuracy compared to general network. An important point here is the specialist networks are only trained with the error cases, which is only 3% of the data in all. It can infer that the specified tasks are easier to be learned. Secondly, we can see that the overall performance of the general network that trained by all the data can reach the accuracy of 82.32%. However, the CNNG after reflection is able to reach 90.88%. The specialist networks are able to have a higher performance in their specific tasks. Therefore, when the task classifier is assigning the task into the specialist networks, it is going to have a higher probability to predict the correct output compared to use the general network. This experiment explains the improved performance of CNNG and provides strong evidence that the reflection algorithm is successfully decomposing the large tasks into easier subtasks.

Discussion and Future Work

This architecture is different from the current trend of deep learning. Instead of blindly adding the layers, we try to expand the network in another way. The combined knowledge in cognitive science and learning science might provide new perspectives to deep learning. Also, this architecture might help to create a deep learning architecture with higher interpretability. The task classifier can allow human to identify which strategy that the learning system is using for a prediction. The combined relatively small networks also lower the difficulty to find an explanation.

For future work, firstly, a more general reflection algorithm and cohesive learning theory should be investigated. We believe that there can be a method to find the best clusters when reflecting the error cases that is able to guarantee the accuracy of task classifier and specialist networks together.

Also, based on our current result, three simple feedforward network can have a better performance compared to a convolutional neural network. It partly shows that using a series of neural network might be able to learn a task with much higher efficiently. We might try to a mathematical proof that adding more networks in CNNG will require much lower amount of training data compared to a deep neural network.

In our implementation of CNNG with reflection currently, we only performed the reflection for once. We are interested in building a multi-layer CNNG with several times of reflection, which will be a hierarchical group of neural networks.

It may be able to improve the performance for a further step.

Conclusion

In this paper, we performed the Collaborative Neural Network Group by reflection. Our method evolves a single neural network to a CNNG by reflecting on its errors. Based on the result of our experiment, the CNNG by reflection could largely improve the accuracy with a low training cost. It is able to reach the accuracy of 99.65% in MNIST, 98.81% in Fashion-MNIST, and 90.88% in EMNIST.

References

[Boud, Keogh, and Walker 2013] Boud, D.; Keogh, R.; and Walker, D. 2013. *Reflection: Turning experience into learning*. Routledge.

[Guangyu Robert Yang 2017] Guangyu Robert Yang, H. Francis Song, W. T. N. X.-J. W. 2017.

[Guerguiev, Lillicrap, and Richards 2017] Guerguiev, J.; Lillicrap, T. P.; and Richards, B. A. 2017. Towards deep learning with segregated dendrites. *eLife* 6.

[Hansen and Salamon 1990] Hansen, L. K., and Salamon, P. 1990. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence* 12(10):993–1001.

[Huang et al. 2017] Huang, G.; Li, Y.; Pleiss, G.; Liu, Z.; Hopcroft, J. E.; and Weinberger, K. Q. 2017. Snapshot ensembles: Train 1, get m for free. *arXiv preprint arXiv:1704.00109*.

[Islam, Yao, and Murase 2003] Islam, M. M.; Yao, X.; and Murase, K. 2003. A constructive algorithm for training cooperative neural network ensembles. *IEEE Transactions on neural networks* 14(4):820–834.

[Jane X Wang 2018] Jane X Wang, Zeb Kurth-Nelson, D. K. D. T.-H. S. J. Z. L. D. H. M. B. 2018. Prefrontal cortex as a meta-reinforcement learning system. *biorxiv preprint https://doi.org/10.1101/295964*.

[Kandel et al. 2000] Kandel, E. R.; Schwartz, J. H.; Jessell, T. M.; of Biochemistry, D.; Jessell, M. B. T.; Siegelbaum, S.; and Hudspeth, A. 2000. *Principles of neural science*, volume 4. McGraw-hill New York.

[Krogh and Vedelsby 1995] Krogh, A., and Vedelsby, J. 1995. Neural network ensembles, cross validation, and active learning. In *Advances in neural information processing systems*, 231–238.

[LeCun, Bengio, and Hinton 2015] LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature* 521(7553):436.

[Liu and Yao 1999] Liu, Y., and Yao, X. 1999. Simultaneous training of negatively correlated neural networks in an ensemble. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29(6):716–725.

[Riding and Sadler-Smith 1997] Riding, R. J., and Sadler-Smith, E. 1997. Cognitive style and learning strategies: Some implications for training design. *International Journal of Training and Development* 1(3):199–208.

[Roelfsema and Holtmaat 2018] Roelfsema, P., and Holtmaat, A. 2018. Control of synaptic plasticity in deep cortical networks. 19:166–180.